Clinical trials to compare two treatments

Papageorgiou, Spyridon N

DOI: https://doi.org/10.1080/14653125.2017.1279465

Originally published at:
Papageorgiou, Spyridon N (2017). Clinical trials to compare two treatments. Journal of Orthodontics, 44(1):59-61.
DOI: https://doi.org/10.1080/14653125.2017.1279465

Clinical trials to compare two treatments

Spyridon N. Papageorgiou

Clinic of Orthodontics and Pediatric Dentistry, Center of Dental Medicine, University of Zurich, Plattenstrasse 11, Zurich 8032, Switzerland

Correspondence to: Spyridon N. Papageorgiou, Clinic of Orthodontics and Pediatric Dentistry, Center of Dental Medicine, University of Zurich, Plattenstrasse 11, Zurich 8032, Switzerland. Tel.:+41 44 634 32 87, Fax: +41 44 634 43 35; E-mail: snpapage@gmail.com.

Manuscript

Orthodontic researchers have developed a novel way to enhance comprehensive orthodontic treatment with fixed appliances using supplemental irradiation with infrared light. Their rationale is founded on basic biological principles and their own clinical experience with prototype appliances. In order to initiate commercial production of infrared irradiation as an adjunct to orthodontic treatment and market their own device under the name *Infrabrace*, they decided to test their experimental appliance with a clinical trial.

The authors performed a controlled clinical trial to assess the treatment effects of *Infrabrace* in a parallel trial with two patient groups. Patients in the first group (henceforth, called the experimental group) were treated with fixed appliances in conjunction with a daily 30 minute use of the experimental *Infrabrace* appliance, while patients in the second group (henceforth, called the control group) were treated with conventional fixed appliances without *Infrabrace*. The authors recruited for this study 60 consecutive adolescent patients from the private practice of an orthodontist, who were divided in two groups of 30 patients each with similar age and sex (Table). Both groups were treated with the same protocol regarding fixed appliance, wire progression, treatment mechanics and intervals between appointments. The primary outcome of the trial was

overall treatment duration from appliance insertion to appliance removal in months and the secondary outcome was an assessment of outcome after treatment with the use of the Peer Assessment Rating (PAR) index (Richmond *et al*. 1992). Measurements were performed by a calibrated external assessor who was not involved in treatment and was blinded to which patient belonged to which group. Data were analysed descriptively with mean and standard deviation (SD) of each group and differences between groups were checked with Student's t-test for independent samples. They found that considerable differences existed in the treatment duration of the experimental group (mean=18.5 months; SD=3.1 months) and the control group (mean=24.9 months; SD=4.6 months), which were statistically significant (P<0.001; Table). Finally, they found no differences in final PAR score of the experimental group (mean=3.1 points; SD=1.4) and the control group (mean=3.6 points; SD=1.7), which was confirmed from the statistical analysis (P>0.05). They concluded that *Infrabrace* is an effective adjunct for reducing treatment time with fixed orthodontic appliances.

Based on the above trial report, which of the following statements, if any, are correct:

(a) The trial robustly assessed the efficacy of *Infrabrace.*

(b) The trial robustly assessed the efficiency of *Infrabrace.*

(c) Trial outcomes were appropriately measured without bias.

(d) Differences in the results of the two treatment groups can be attributed to *Infrabrace*.

Answers

Statement (c) is true; statements (a), (b), and (d) are false.

(a) The trial robustly assessed the efficacy of *Infrabrace*. Efficacy in the present example would be defined as the extent to which *Infrabrace* produces a beneficial result under ideal conditions (Porta, 2014). The authors measured the occlusal outcome of treatment with PAR scoring and

found that no significant differences existed between the two groups after treatment ($P>0.05$). Although it cannot be formally concluded that no difference exists in the PAR scores of the two groups (since an absence of evidence is not evidence of absence), we can be confident that the final PAR scores of the two groups seem very similar. The PAR index can be used to measure both the severity of a malocclusion and the outcome of orthodontic treatment (Richmond et al., 1992). However, no information is given about the initial PAR scores of the two groups, a fact that does not let us assess, if equally "difficult" cases were included in each group. A better alternative would be to also measure the baseline PAR scores of each case prior to treatment and incorporate this in the trial, for example by calculating the absolute or relative PAR change for each patient through treatment. We can conclude for the present trial that both the experimental and the control protocols can be used to treat to a similar standard (since the final PAR scores of the two groups were similar), but we cannot draw any conclusions about the efficacy of *Infrabrace* (since we do not know if the PAR reduction of the two groups were different).

(b) The trial robustly assessed the efficiency of *Infrabrace*. Efficiency pertains to the effects or end results achieved in relation to the effort expended in terms of money, resources, and time (Porta, 2014). In the present example, if the two groups show similar treatment effects, but the duration of treatment is drastically reduced in the *Infrabrace* group, it then follows that treatment efficiency is improved. Following from (a), it is clear that this statement is only partly true, since the authors did not robustly assess the efficacy of treatment with *Infrabrace*. It would be more appropriate to say, that if the reduction in PAR score was similar in the two groups, but the time needed to finish treatment is reduced with the use of *Infrabrace*, then treatment efficiency would be improved. Another alternative would be to calculate a composite outcome measurement of PAR reduction divided by duration, but this might make the trial results and their interpretation more complex. Additionally, any improvement in treatment effects or reduction in treatment duration should ideally

be viewed together with any additional costs of treatment and side effects that might be associated with *Infrabrace*.

(c) Trial outcomes were appropriately measured without bias. The authors used an external assessor to measure all trial outcomes, which minimised detection bias. It is known that measuring the results of compared interventions can be influenced by a number of factors, including an assessor's personal preferences or present expectations from a trial that has cost plenty of time and money to the researchers. As the outcome assessor was blinded, the influence from personal preferences or expectations is minimised, while measurement errors are appropriately handled by the *a priori* calibration and *post hoc* analytic techniques (Bland and Altman, 1986).

(d) Differences in the results of the two treatment groups can be attributed to *Infrabrace*. This statement is false, since the existence of bias cannot be safely ruled out from the present trial. True baseline equivalence of experimental and the control groups prior to treatment has not been ascertained in the present trial. The outcome of orthodontic treatment might be influenced by several known or unknown factors, such as malocclusion severity, patient compliance, the genetically-determined biological response of each patient, systematic diseases, smoking, and interactions with any medications taken. Ideally, *a priori* equivalence would be attained through random patient allocation in the experimental and control groups, which would ensure that the distribution of all known or unknown confounding factors is similar between groups. In this case, any differences found between the experimental and the control group could be attributed to the supplemental use of *Infrabrace*. In cases where randomisation is not feasible, statistical methods that mimic randomisation like propensity score matching can also be used, although they are more complex and not without criticism (Stuart, 2010). Therefore, randomisation is the best approach to ensure baseline equivalence in clinical trials of comparative effectiveness.

References

Richmond S, Shaw WC, Roberts CT, Andrews M. The PAR Index (Peer Assessment Rating): methods to determine outcome of orthodontic treatment in terms of improvement and standards. Eur J Orthod. 1992 Jun;14(3):180-7.

Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010 Feb 1;25(1):1-21.

Porta M. A Dictionary of Epidemiology. 6th ed. New York, NY: Oxford University Press; 2014.

Table. Patient characteristics and outcomes of the given trial example.

		Experimental	Control	P value
Patient characteristics				
	Patients – n	30	30	
	Male/female – n	17/13	14/16	
	Age in years – mean (SD)	13.8 (1.5)	14.1 (2.0)	
Outcomes				
	PAR score after treatment– mean (SD)	3.1 (1.4)	3.6 (1.7)	0.187
	Treatment duration in years – mean (SD)	18.5 (3.1)	24.9 (4.6)	<0.001

SD, standard deviation; PAR, peer assessment rating.

Appendix. Dataset from the given trial example.

Nr	groupcode	group	txduration	par1	par2	sex	age
1	1	experimental	18.57	28	8.00	1	12
2	1	experimental	15.33	25	3.00	1	12
3	1	experimental	16.43	26	2.00	0	16
4	1	experimental	20.07	24	4.00	1	17
5	0	control	23	18	2.00	0	15
6	1	experimental	22.1	28	3.00	0	14
7	1	experimental	18.47	36	2.00	0	13
8	1	experimental	16	22	3.00	1	13
9	1	experimental	15.27	24	3.00	1	14
10	0	control	19.4	33	2.00	1	12
11	0	control	27.4	24	4.00	1	12
12	1	experimental	24.3	33	2.00	0	14
13	1	experimental	16.63	24	2.00	0	13
14	0	control	28	52	2.00	1	14
15	1	experimental	17.63	36	2.00	1	14
16	1	experimental	18.47	29	2.00	1	13
17	0	control	18.73	25	4.00	1	14
18	1	experimental	18	21	3.00	0	16
19	1	experimental	16.3	25	2.00	0	14
20	1	experimental	15.4	30	3.00	0	16
21	1	experimental	16.53	23	4.00	1	15
22	0	control	26.6	53	11.00	0	13
23	0	control	22.9	29	2.00	0	12
24	0	control	29.5	53	2.00	0	13
25	0	control	29.7	23	4.00	1	12
26	1	experimental	16	30	2.00	0	12
27	0	control	25.57	24	3.00	1	13
28	0	control	37.23	43	2.00	1	15
29	0	control	20.63	33	5.00	0	15
30	1	experimental	16.33	21	2.00	0	13
31	0	control	22.73	49	2.00	0	13
32	1	experimental	16	25	2.00	1	12
33	0	control	25.7	20	3.00	0	14
34	0	control	25.23	29	3.00	1	16
35	1	experimental	16.73	32	6.00	1	15
36	0	control	20.77	47	9.00	1	19
37	1	experimental	15	16	2.00	1	12
38	0	control	21.93	47	2.00	0	19
39	0	control	31.03	27	5.00	0	15
40	0	control	24.2	32	2.00	0	14
41	0	control	22.17	31	3.00	0	12
42	1	experimental	17	18	3.00	0	13
43	0	control	25.2	56	7.00	1	14
44	0	control	26.4	46	4.00	0	15

45	1	experimental	18	16	4.00	1	15
46	0	control	26.77	51	19.00	1	12
47	0	control	36.5	30	4.00	0	12
48	1	experimental	25.56	49	3.00	1	13
49	0	control	19	14	4.00	1	14
50	1	experimental	20.27	36	2.00	1	12
51	1	experimental	23.9	35	5.00	1	16
52	1	experimental	19.17	21	2.00	0	13
53	1	experimental	17.03	25	4.00	0	15
54	0	control	22.47	31	2.00	1	13
55	0	control	24.37	39	6.00	0	14
56	0	control	23	28	2.00	1	12
57	1	experimental	22.07	39	4.00	1	13
58	0	control	23.57	41	11.00	1	19
59	0	control	24.47	27	6.00	0	15
60	0	control	18.57	19	6.00	0	13